SEASONAL REVENUE FORECASTING AND RISK MANAGEMENT

Andrew Shi, *Diamond Bar High School, 21400 Pathfinder Rd, Diamond Bar, CA 91765, 909-656-3424, andrewpianiluv@gmail.com*
Cosmin Borsa, *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, 805-893-2129, cosmin_borsa@umail.ucsb.edu*

ABSTRACT

Revenue forecasting is critical for companies to make data-driven decisions, especially for those that exhibit revenue seasonality. The purpose of this study is to build a robust time series model to forecast seasonal revenue and perform risk analysis. Using Booking Holdings as an example, we develop a Seasonal Autoregressive Integrated Moving Average (SARIMA) model by transforming and differencing the time series data. We then estimate the optimal parameters of the SARIMA model and conduct rigorous diagnostic tests. Finally, we apply the proposed SARIMA model to Booking Holdings' financial data to analyze its liquidity and solvency risk under the impacts of the COVID-19 pandemic.

Keywords: SARIMA model, time series forecasting, COVID-19 pandemic, risk analysis, model validation

INTRODUCTION

Revenue forecasting derives much of its applications from using historical data to predict future revenue. It is the most widely followed performance metric by financial analysts and plays an important role in the valuation of a firm [1]. An accurate revenue forecast allows firms to properly allocate company resources and strategically make investment, operational, and financing decisions. One of the major challenges for accurate revenue forecasting is seasonal variation. For example, the online travel agency (OTA) industry features strong seasonality across the board, making its revenue forecasting less accurate and more challenging. Previous studies show that revenue seasonality significantly influences the impact of earnings announcements on stock price performances [2].

Several statistical models have been developed to forecast time series data in general and revenue in particular, such as the exponential smoothing model, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, and Autoregressive Integrated Moving Average (ARIMA) model. Although these models generally are able to catch the trend within time series data, they do not perform well in a time series with seasonal variations. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model, an extension of the ARIMA model, is "capable of modelling a wide range of seasonal data" [3].

In this paper, we aim to build a robust revenue forecasting model for Booking Holdings (BKNG), the world's largest OTA by revenue. The purpose is to better understand BNKG's seasonal revenue variations, especially under the influence of the unprecedented COVID-19 pandemic. More specifically, we use BKNG's quarterly revenue data to create a SARIMA model through the Box-Cox transformation and differentiation techniques. We then estimate the optimal parameters of the SARIMA model and perform model validation by comparing the differences between forecasted and actual revenues in both pre- and post-pandemic periods. Lastly, we apply the proposed SARIMA model to forecast BKNG's quarterly revenues and analyze its liquidity and solvency risk in 2020.

The revenue data in this study is obtained from The Macrotrends LLC, a platform that records over 50 years of financial revenue data [4]. The statistical analysis part of this paper is written by a popular statistical software, R.

DATA TRANSFORMATION

In order to build a SARIMA model for BKNG's quarterly revenue data, a series of transformations are performed to make the data stationary. First, we perform a Box-Cox transformation to stabilize the variance of our time series. We then differentiate at lag-4 to remove seasonality and at lag-1 to remove linear trend.

In Figure 1, we decompose BKNG's quarterly revenue data into three components: trend, seasonal, and random. It is observed that the variance of the random component is unstable. To stabilize variance and transform the data to normal distribution, we perform a Box-Cox transformation using Equation 1.

$$Y_t = \frac{1}{\lambda}\left(X_t^{\lambda} - 1\right) \qquad (1)$$

Since the revenue data set is collected on a quarterly frequency, we then difference the data at lag-4 using Equation 2 to further remove the observed seasonality.

$$Z_t = Y_t - Y_{t-4} \qquad (2)$$

In order to remove the existing linear trend, we difference again at lag-1 by using Equation 3.

$$W_t = Z_t - Z_{t-1} \qquad (3)$$

The resulting time series data is shown in Figure 2, in which the linear trend is no longer observed and the variance of the time series decreases by 57%. We conclude that the differentiations at lag-4 and lag-1 are necessary because of the decrease in variance and the removal of the seasonal and linear trends.

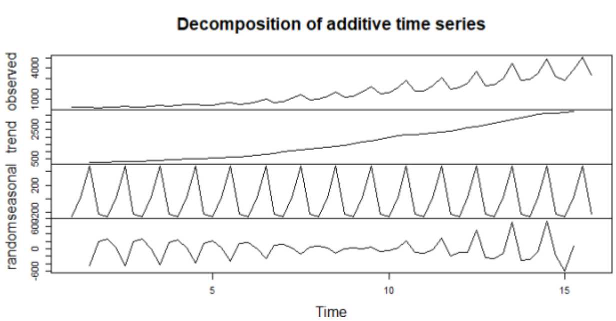

FIG. 1. Decomposition of the time series data set into the trend,seasonal, and random components.

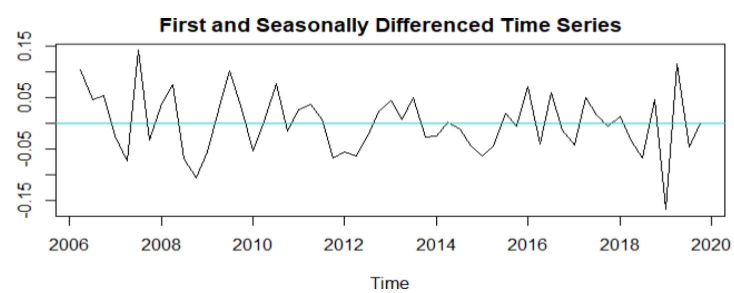

FIG. 2. Time series data differenced at lag-4 and lag-1 (the seasonal and linear trends are no longer apparent)

MODEL CONSTRUCTION

The SARIMA Model

The SARIMA model is specified by as $SARIMA(p, d, q)(P, D, Q)_S$ [5], where parameters *p* and *q* represent the orders of the non-seasonal autoregressive (AR) and moving average (MA) models, respectively. *P* and *Q* represent the orders of the seasonal AR and MA models, respectively. *d* and *D* are the non-seasonal and seasonal differencing parameters, respectively. The parameter, *S*, represents the seasonal frequency of the data. Since quarterly revenue data is used in this research, the *S* value is set at 4.

Estimating Parameters for the SARIMA Model

The optimal parameters of the proposed SARIMA model are estimated based on the transformed data. We examine the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots of the seasonally-differenced time series. In Figure 3 and 4, we find that there are a few lags outside of the confidence intervals, which are denoted by the dotted blue lines. These plots reasonably suggest that the order of p and P are botheither 0 or 1. The order of q is 0 and the order of Q is either 1 or 2. Since we difference seasonally at lag-4 and non-seasonally at lag-1, the order of *d* and *D* are both 1.

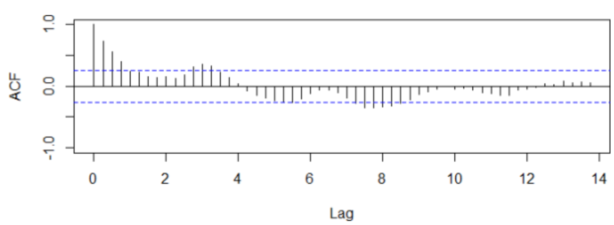

FIG. 3. ACF plot of the seasonally differenced time series

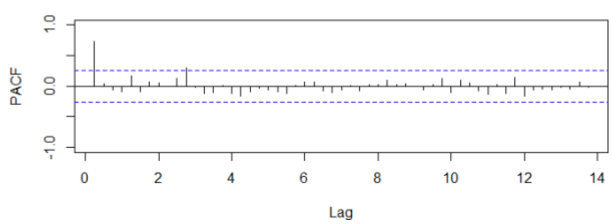

FIG. 4. PACF plot of the seasonally differenced time series.

Building the Model

Using the suggested parameters for the SARIMA model, we construct a list of 8 model candidates shown in Appendix 1. We evaluate the Akaike information criterion corrected for bias (AICc) for each of these models by using Equation 4. The model with the lowest AICc value will be considered the best fit.

$$AICc = -\log L\left(\theta_q, \phi_p, S\left(\theta_q, \phi_q\right)/n\right) + \frac{2(p+q+1)\,n}{n-p-q-2} \qquad (4)$$

The AICc values for each of our models are shown in Appendix 1 as well. We find that Model 5, $SARIMA(0,1,0)(0,1,2)_4$,has the lowest AICc value, which makes it the best fit for our data.

Diagnostic Testing

Several diagnostic tests are performed on $SARIMA(0,1,0)(0,1,2)_4$ to further explore its attributes.

Figure 5 depicts the plot and histogram of the residuals of the proposed SARIMA model. The plot resembles that of white noise and the histogram is somewhat normally distributed. Figure 6 displays the Q-Q plot for the proposed SARIMA model. Clear departures from the Q-Q line are found in the graph, indicating a heavy-tailed distribution, which is common in time series data. Figure 7 shows ACF and PACF plots of the residuals of $SARIMA(0,1,0)(0,1,2)_4$. It is found that all spikes are inside the confidence interval and resemble white noise. The spike at lag-0 for the ACF plot always has a value of 1.

Last, we perform four Portmanteau tests on our model. The Shapiro-Wilk test examines the normality of the residuals, the Box-Pierce and Ljung-Box tests assess the independence of the residuals, and the McLeod-Li test checks for non-linear dependence of the residuals. The results of tests are shown in Table 1. The *p-values* of all four tests exceed 0.05, suggesting that our residuals are independent of each other and are representative of white noise.

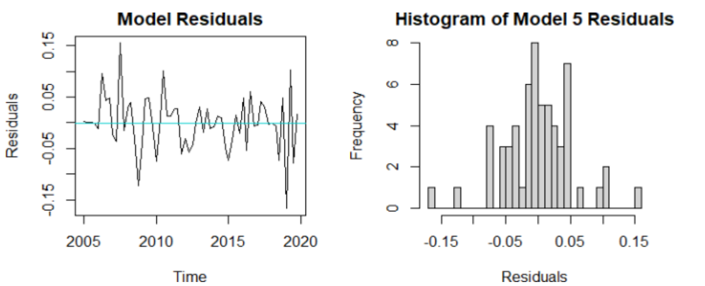

FIG. 5. Plot and histogram of $SARIMA(0,1,0)(0,1,2)_4$ residuals

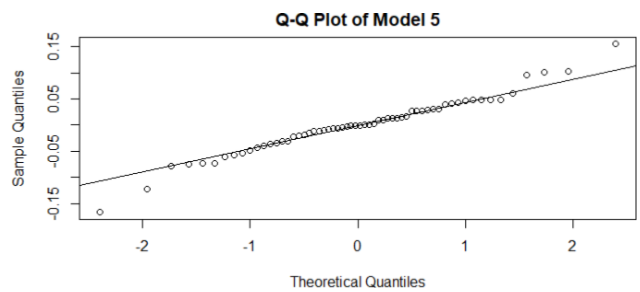

FIG. 6. Q-Q plot of $SARIMA(0,1,0)(0,1,2)_4$.

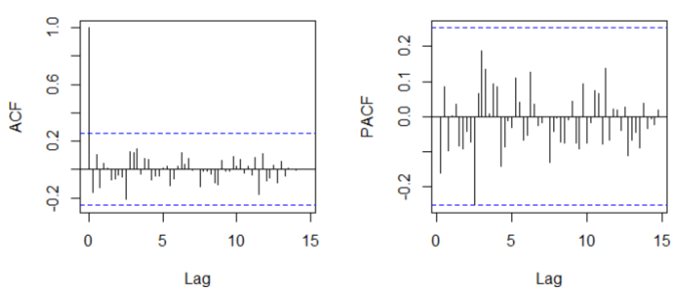

FIG. 7. ACF and PACF plots of the residuals of SARIMA(0,1,0)(0,1,2)$_4$.

Tests	*p-value*
Shapiro-Wilk Test	0.1673318
Box-Pierce Test	0.5471178
Ljung-Box Test	0.3851826
McLeod-Li Test	0.6078178

Table 1. p-values for four Portmanteau diagnostic tests.

FORECASTING AND RISK ANALYSIS

In this section, we forecast BKNG's revenue based on the proposed model, $SARIMA(0,1,0)(0,1,2)_4$. Figure 8 shows BKNG's forecasted quarterly revenue in 2020. The dotted green lines define the 95% prediction interval of the projected revenue, assuming no impact of COVID-19 on sales. The lower blue line represents the actual revenue of BKNG in 2020. With the forecasted total revenue, we can not only project other income statement and balance sheet items based on their historical proportional relationships with total revenue, but also analyze the profitability, liquidity, and risk of bankruptcy for

BKNG. A scenario analysis is also conducted to examine the impact of the COVID-19 pandemic on BKNG's operating income in the best, base, and worst scenarios.

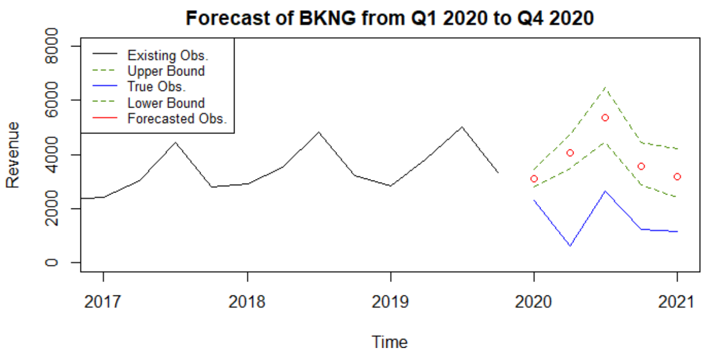

FIG. 8. Graph of the forecasted revenue for BKNG in 2020.

Income Statement Analysis

Based on BKNG's historical income statements from 2017 to 2019, we identify expense items that vary proportional to total revenue and items that do not vary proportional to total revenue. For example, we find that marketing expenses vary proportionally, around 33%, to thetotal revenue, while other expenses within the Selling, General and Administrative (SG&A) expenses remain relatively fixed at around $4 billion per year. Using such methods, we can use BKNG's total revenue to further analyze its operating income and net profit.

In order for BKNG to make positive net profit, the company at least needs to be able to cover interest expenses with the operating income. We estimate the interest expenses for BKNG to be around $266 million based on its debt level. To achieve operating income of more than $266 million, BKNG needs a minimum total revenue of around $2.5 billion. Since our 95% confidence interval projects an annual total revenue between $13.62 billion and $19.15 billion, the probability of BKNG's annual total revenue falling below $2.5 billion is significantly less than 2.5% ((1-95%)/2). The income statement analysis suggests that BKNG's forecasted revenue is sufficient to cover various expenses and meet its interest expense obligations, indicating low solvency risk.

Balance Sheet Analysis

Aside from interest expenses, BKNG has to pay off current liabilities every year. Using a method similar to the income statement analysis, we estimate the liabilities and assets of BKNG in 2020 under normal circumstances based on their historical relationship with total revenue. We conservatively project that the growth of the company's current liabilities and current assets both increase by 5% every year. Subtracting the current liabilities from the current assets, we project that the company produces a surplus of $3.85 billion working capital every year. The robust liquidity position indicates BKNG's strong ability to meet its current debt obligations and short-term cash requirements.

Both the income statement and balance sheet analysis suggest that BKNG is in a secure financial position and should have minimal risk of bankruptcy.

COVID-19 Scenario Analysis

In this section, we present the best-case, base-case, and worst-case scenarios of BKNG's projected revenue using $SARIMA(0,1,0)(0,1,2)_4$ and calculate the impact of COVID-19 on operating income in the three scenarios. The best-case scenario projects a revenue of $19.15 billion and a total operating income of $7.86 billion. Compared to its actual operating loss of $0.63 billion in 2020, we estimate that the pandemic caused a reduction of $8.49 billion operating income in the best-case scenario. The base-case scenario projects a revenue of $16.14 billion and an operating income of $5.84 billion. Compared to the actual operating income, the impact of the pandemic was a $6.47 billion loss in operating income. Lastly, the worst-case scenario projects a revenue of $13.62 billion and an operating income of $4.15 billion, suggesting a $4.78 billion reduction in operating income due to the pandemic.

To further understand the forecasting gaps between our SARIMA model and BKNG's actual 2020 revenue, we compare the forecasting errors of 2020 with that of 2019, a pre-pandemic year. In 2019, under a business-as-usual assumption, the Mean Absolute Percent Error (MAPE) is low, ranging from 5.91% to 9.72%, as shown in Table 2. However, the errors rise substantially to 39.43% to 582.45% in 2020. The results indicate that the proposed SARIMA model is quite accurate in revenue forecasting under normal circumstances and that the large forecasting errors of 2020 are mainly attributed to the pandemic.

Quarter	*Actual Value(in millions)*	*Forecasted Value (in millions)*	*MAPE**
Q1 2019	$2,837.00	$3,112.74	9.72%
Q2 2019	$3,850.00	$4077.47	5.91%
Q3 2019	$5,040.00	$5366.87	6.49%
Q4 2019	$3,339.00	$3583.69	7.33%
Q1 2020	$2,288.00	$3190.10	39.43%
Q2 2020	$630.00	$4299.46	582.45%
Q3 2020	$1,640.00	$5537.06	109.74%
Q4 2020	$1,238.00	$3713.97	200.00%

Table 2. Forecasting accuracy of 2019 and 2020 revenue

CONCLUSION

In this project, we model BKNG's revenue data with a SARIMA model. BKNG's historical revenue data exhibits steady growth and seasonality over time. Nevertheless, the proposed SARIMA model shows decent accuracy in seasonal revenue forecasting, enabling us to not only perform analyses on the company's profitability, liquidity, and solvency risk, but also conduct a scenario analysis on the impact of the COVID-19 pandemic on BKNG's operating income. Given our projection, BKNG is in a strong liquidity and solvency position with little bankruptcy risk in the near future.

A limitation of our research is that the forecasting error of the SARIMA model is high for long-tail events, such as the COVID-19 pandemic. In the future, our work can be extended by incorporating machine learning algorithms and other statistical models to improve the accuracy of our revenue forecast. If mature, these techniques may become future common practices in business analytics.

REFERENCES

[1] Rees, Lynn L. & Sivaramakrishnan, K. (2007). The effect of meeting or beating revenue forecasts on the association between quarterly returns and earnings forecast errors, *Contemporary Accounting Research*, 24(1), 259–290e

[2] Salamon, Gerald L. & Stober , T. L. (1994). Cross-quarter differences in stock price responses to earnings announcements: fourth-quarter and seasonality influences, *Contemporary Accounting Research*, 11(1), 297–330.

[3] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. (OTexts.com/fpp3. Accessed on 7/11/2021).

[4] Booking Holdings Revenue 2006-2021, Macrotrends. Accessed June 23, 2021. https://www.macrotrends.net/stocks/charts/BKNG/booking-holdings/revenue

[5] Brockwell, P. J. & Davis, R. A. Introduction to Time Series and Forecasting. Cham: Springer, 2016.

APPENDIX

Appendix 1:

Model Candidates	AICc Values of Models
Model 1: SARIMA$(0,1,0)(0,1,1)_4$	-154.893
Model 2: SARIMA$(1,1,0)(0,1,1)_4$	-153.9033
Model 3: SARIMA$(0,1,0)(1,1,1)_4$	-153.9203
Model 4: SARIMA$(1,1,0)(1,1,1)_4$	-152.3361
Model 5: SARIMA$(0,1,0)(0,1,2)_4$	-155.1404
Model 6: SARIMA$(1,1,0)(0,1,2)_4$	-154.3354
Model 7: SARIMA$(0,1,0)(1,1,2)_4$	-154.172
Model 8: SARIMA$(1,1,0)(1,1,2)_4$	-153.1587